

7 Collate information on the quality status of age reading at MS institutes (ToR g)

Quality Status Of Age Reading At MS Institutes.

The table (Quality Status Of Age Reading At Institutes.xls) is downloadable separately.

7.1 Internal Quality Management

Three categories of internal quality management for age reading have been identified:

- Quality of the age readings is managed by an individual age reader. CS (calcified structures) preparation and age reading procedures and manuals may not exist for all stocks/species. Typically there is only one age reader for each stock/species and there are no bilateral agreements with readers at other institutes. After an interval of time has passed since the initial readings (usually a minimum of several weeks), the reader re-reads a percentage of each sample of calcified structures. If the readings for an individual CS do not agree and the discrepancy cannot be resolved, that CS is excluded from the assessment. Original CS may not be catalogued and stored in a way that will prevent damage, loss or deterioration. This is the least preferable level of internal quality management, as calcified structures and images can be subconsciously memorised with increasing use, leading to the development of unintentional bias. This situation can be improved by the establishment of reference collections and including some samples from a reference collection in each sample of CS that are re-read, by setting up bilateral agreements and by participation in relevant age calibration exchanges (Ex) and age calibration workshops (WKAC). It is important that reference collections are sufficiently large to avoid frequent use of individual images.
- Quality of the age readings is managed by two or more readers. CS preparation and age reading procedures and manuals exist for all stocks/species. At intervals, each reader re-reads a percentage of the Calcified Structures (CS) read by the other reader(s). If the readings for an individual CS do not agree and the discrepancy cannot be resolved, that CS is excluded from the assessment. There may be bilateral agreements with readers at other institutes. The number of age readers may be low and it is difficult to obtain meaningful results from the analysis of age reader bias. Original CS are catalogued and stored in a way that will prevent damage, loss or deterioration. There is some participation in relevant age calibration exchanges (Ex) and age calibration workshops (WKAC). This situation can be improved by the establishment of reference collections and including some samples from a reference collection in each sample of CS that are re-read, by setting up bilateral agreements with other institutes and increased participation in relevant age calibration exchanges (Ex) and age calibration workshops (WKAC). The management of all documents is controlled and properly managed (unauthorised changes to quality documents are not permitted). It is likely that institutes that meet these requirements may have, or be capable of obtaining quality management certification, e.g. ISO 9001: 2008.

- There are several age readers for each stock/species and bilateral agreements exist with age readers at other institutes who read CS from the same stock/species. CS preparation and age reading procedures and manuals exist for all stocks/species. Reference collections are used for each stock/species and these are updated at intervals with new CS, to reduce the possibility of unintentional bias caused by too frequent exposure to the same CS or (CS image). There are regular scheduled exercises to monitor changes in the age reader performance. There are specified training programmes that require the trainee to be closely supervised for a period of time, before they can enter the cycle of quality management for age readers who contribute data to assessments. A quality manual exists and this defines how reference collections are managed, how training programmes are managed and how quality checks and controls are carried out. The age readers take part in all relevant age calibration exchanges (EX) and age calibration workshops (WKAC). This is the most preferable level of internal quality management. Original CS are catalogued and stored in a way that will prevent damage, loss or deterioration. The reference collections are managed using an image database (N.B. WebGR can be installed locally for this purpose). The management of all documents is controlled and is properly managed, typically by means of an electronic document repository. Unauthorised changes to quality documents are not permitted and printed versions will be watermarked as uncontrolled documents. It is likely that Institutes that meet these requirements may have, or be capable of obtaining quality management certification, e.g. ISO 9001: 2008 and ISO 17025 (accreditation).

7.2 External Quality Management

Accreditation is the highest level of external quality management and two MS institutes currently hold national accreditation, CEFAS (ISO 17025 - UKAS) and IVLO (ISO 17025 - BELAC).

7.3 Study on Harmonised methodology of age estimation Procedures

Most of European fish stocks are assessed using age-based models. In this context, the quality of these data plays a vital role in management of fish resources. Errors in age estimation can be caused by accuracy and/or precision issues (Campana, 2001) and they have to be detected and quantified. Accuracy refers to the closeness between measurements and their true value. Precision is defined as the variability in the age readings. Within and between age reading laboratories there will inevitably be disagreements between age readers.

The European Commission supported the development of a European Fish Ageing Network with two Concerted Actions, i.e. EFAN (European Fish Aging Network) and TACADAR (Towards Accreditation and Certification of Age Determination of Aquatic Resources) from 1997 to 2006. The overall theme of both CAs was to harmonise the fish age estimation, mostly carried out by interpreting calcified structures.

Since 2006, the ICES Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS) organises each year of calibration exercises and workshops on calcified structures to estimate precision and relative/absolute bias in the age estimations from readers based in different ageing laboratories.

In 2007, the PGCCDBS meeting decided to produce the first international age length key (ALK) for the sole in the Eastern English Channel (ICES area VIIId) in 2008. In the future, the international ALK should develop requiring the harmonisation of the age estimation procedures.

Proposition:

The objective for this proposal is to develop a harmonised methodology and quality assurance process for age estimation procedures for a small region with a group of countries, before trying to achieve general harmonisation of these procedures for all ICES areas. Firstly, we need to identify the region and common species/stocks to be used in this trial harmonisation. The techniques for each step from sampling of the calcified pieces to the storage/archiving then need to be described and compared per country and per species/stocks. Common tools for quality indication and assurance need to be developed for each step of the age estimation procedure. The most qualitative procedures will then be used for optimisation and harmonisation of the protocols (methods of preparation, material etc.) per species/stock and among species/stocks. This will then result in a harmonised quality control mechanism and the development of guidelines for good techniques and good operating procedures. According to these harmonised protocols, we will also need to develop consistent training for new age readers (per species/stock). These new sets of harmonised protocols will give more strength for accreditation processes and will strengthen already accredited labs.

This study was presented during the Regional Coordination Meeting for the North Atlantic (RCM NA) 2010 and during the ICES Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS) 2011 meeting.

WKNARC is supporting the idea of this study with a small number of case studies.

7.4 Proposal for the use of a standard grading system by age readers of their own readings to register the confidence level the reader has in their otolith readings, reflecting the quality of the data.

PGCCDBS recommends the use of a standard grading system by the mackerel age reader of his/her own readings (e.g. high, medium, low) be considered during the WKNARC as a standard that could be applied in all age calibration ex-changes and/or WKs (ref. WKARMAC 2010).

It is a recommendation of the WKARMAC 2010 Workshop to register the confidence level the reader has in their otolith reading, reflecting the quality of the data. Most readers should use a scale of 3 levels of quality:

- Rings can be counted with certainty: 1
- Rings can be counted, but with difficulty and some doubt: 2
- Rings cannot be counted, the otolith is unreadable: 3

WKNARC supports the PGCCDBS recommendation for the use of a grading system by the age reader of their own readings in all age calibration exchanges and/or WKs.

Reading the ages of calcified structures is a subjective process, where the reader uses their acquired knowledge of CS growth zone patterns and the biology of a species-stock, to estimate the age of individual CS that are believed to be from that species-stock. Typically, age readers who provide the age data that are used in stock assessments are experienced in reading the CS of the species-stocks studied at their institutes. Several different scales of different lengths have been used at MS institutes and

currently, 3-point and 4-point scales are used to allow the age readers to record their confidence in their age readings of individual calcified structures. The points on the quality scale of the readings should refer to an age interpretation protocol for the species-stock, to ensure that common principles are used to define the quality of the age readings and minimise the subjective nature of individual age reader quality judgments.

The idea of a grading scale is important, as the readers own assessment of the quality of the ageing structure may well be associated with a bias in interpretation results. E.g. it may well be that 'poor otoliths' are simply otoliths from e.g. slow-growing individuals of the stock, or are from specific parts of the distribution area. Growth is very variable in fish, and age readers should expect variety in the patterns that are used to interpret the age of the calcified structures. It is important to be able to identify such potential bias in the age data. There are many biological, physiological and environmental influences on the growth and some species-stocks may have few calcified structures that are easy to interpret. The use of lowest quality scale value would enable the age reader to give an age estimate, even when the age reader considers the pattern within an ageing structure as poor, knowing that this data will be treated with caution. This in turn enables a revaluation of the poor otoliths and an investigation of which structures the readers should count. The age distribution of the given year/sample can then be updated accordingly and the age reading protocol revised. There is also the risk of introducing bias if only the "good" readings are considered because only a portion of the population is considered, thus the end-users need to be aware of the grading system and what it represents. 'Bad' readings should not be ignored, but it should be examined if a particular pattern lies behind the assignment of 'bad' (growth, migration, stock, etc).

While in most cases the availability of validated CS should eventually increase the readers' confidence in their age readings, there is a lack of validation for many species-stocks. Age data is provided by different countries. Stock assessors and others need an indication of the confidence level that can be applied to the age data that is used to provide a diagnosis of the status of a species-stock. Therefore it is very necessary to have a scale indicating data quality (confidence in the assigned age data), that will enable a common interpretation of the reliability of the age data. Any scale of data quality should take on board both the inherent uncertainty in all age interpretations, as well as the understanding that the individual sense of certainty is related to a given ageing protocol or reading rules that may well be changed should validated calcified structures become more widely available.

The difficulty in interpreting growth zone patterns in CS varies from species to species and from stock to stock. Growth zone features that can be seen relatively clearly in the calcified structures of one species-stock may be very difficult to observe in another species-stock. WKNARC considers that a simple 3-level scale captures the important information.

A major advantage of the 3-point scale is that it can be linked to the PGCCDBS definitions of ageing performance. These are bad ageing performance, where there are serious concerns about the reliability of the age data and/or its value to stock assessment WGs; medium ageing performance, where the age data is sufficiently reliable to be used for stock assessment purposes but improvement is required and good ageing performance, where the age data is considered reliable (see section 4.4 of this report for detailed descriptions of ageing performance.)

Using the same 3-point scale for all species-stocks would standardise the quality values so that data users could eventually rely on the quality value attached to the age data, without having to consider the precise nature of the quality assessments used to derive the quality value. The three scale points can be defined as easy to age with high precision, difficult to age with age with acceptable precision and unreadable or very difficult to age with acceptable precision. Terms like poor, moderate and good quality and/or values relating to the actual growth pattern elements can be used. E.g. clear scale winter rings or otolith translucent annuli that are easily identified, with no unexpected or difficult to resolve growth pattern elements present on the CS. Unreadable CS are classified as difficult to age, as age data from both of these types of CS have the potential to have a detrimental effect on an assessment, if concerns about the reliability of the age data and/or its value to stock assessment WG, are not clearly identified.

WKNARC recommends that Age Calibration Workshops (WKACs), derive descriptors for the three scale points that are applicable to their species-stocks. These can include examples and detailed definitions for these indicators of age reading quality, including quality of the calcified structure and ease of interpretation of the structure.

WKNARC recommends that the following 3-point scale of age reading quality be used by all age readers who provide age data for stock assessments.

AQ1: Easy to age with high precision.

If a scale of 1-100 is applied, where 100 is when the reader has the highest possible confidence in the age reading and 1 is when the reader has no confidence in the age reading, age quality 1 (AQ1), will apply to approximately the top 25 % of the possible quality ratings. AQ1 is an indication that the age data is considered reliable for stock assessment.

AQ2: Difficult to age with age with acceptable precision.

Age quality 2 (AQ2), will apply approximately to age readings within 25 and 75 percentiles of the possible quality ratings. AQ2 is an indication that the age data is sufficiently reliable to be used for stock assessment purposes but improvement is required.

AQ3: Unreadable or very difficult to age with acceptable precision.

Age quality 3 (AQ3), will apply to approximately the lowest 25 % of the possible quality ratings. 3 AQ3 is an indication that there are serious concerns about the reliability of the age data and/or its value to stock assessment WGs.

7.5 Reference

Campana, S., 2001. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology*, 59: 197–242.